Nonparametric Estimation in a Stochastic Volatility Model

Jürgen Franke* Wolfgang Härdle† Jens-Peter Kreiss‡

March 31, 1998

Abstract

In this paper we derive nonparametric stochastic volatility models in discrete time. These models generalize parametric autoregressive random variance models, which have been applied quite successfully to financial time series. For the proposed models we investigate nonparametric kernel smoothers. It is seen that so-called nonparametric deconvolution estimators could be applied in this situation and that consistency results known for nonparametric errors-in-variables models carry over to the situation considered herein.

1 Introduction

Many methods of financial engineering like option pricing or portfolio management crucially depend on the stochastic model of the underlying asset. If $S(t)$ denotes the stock price at time t, then, e.g., the Black-Scholes approach to option pricing is based on modelling $\log S(t)$ as a Wiener process with drift μ and diffusion coefficient or volatility σ :

$$d(\log S(t)) = \alpha\, dt + \sigma dW(t)$$

where $W(t)$ is a standard Wiener process. This particular model is known to be inappropriate in various circumstances. For instance, σ can no longer be assumed to be constant if the time up to exercising the option is rather short. Replacing the constant σ by a positive stochastic process $\sigma(t)$ we arrive at the following equation for the asset price:

$$d(\log S(t)) = \alpha\, dt + \sigma(t) dW(t). \tag{1.1}$$

In the literature, several specific parametric models for the stochastic volatility $\sigma(t)$ have been proposed and used for option pricing. Here, we restrict ourselves to models

*Universität Kaiserslautern, Erwin-Schrödinger-Straße, D-67663 Kaiserslautern, Germany
†Humboldt-Universität zu Berlin, Spandauer Straße 1, D-10178 Berlin, Germany
‡Technische Universität Braunschweig, Pockelsstraße 14, D-38106 Braunschweig, Germany

which characterize $\sigma(t)$ as the solution of a stochastic differential equation for $\log \sigma(t)$ known up to a few parameters. An example is the equation

$$d(\log \sigma(t)) = \lambda(\kappa - \log \sigma(t))dt + \gamma dW^*(t) \tag{1.2}$$

considered by Scott (1987, 1991), Wiggins (1987) and Chesney and Scott (1989). Here, $W^*(t)$ is another standard Wiener process correlated with $W(t)$ of (1.1)

$$dW(t)\ dW^*(t) = \rho\, dt,$$

and α of (1.1), λ, κ, γ and ρ are the unknown model parameters. Other models of a similar structure have been proposed in the literature.

To help to answer the question which stochastic volatility model is appropriate for a particular data set we consider a rather general type of model avoiding the assumption of a particular parametric form of the equation defining $\sigma(t)$. At the beginning, we discretize time, as is also frequently done for parametric models for the purpose of estimating the model parameters. The log-volatility will then satisfy a general nonlinear stochastic difference equation or nonlinear autoregressive scheme. As $\sigma(t)$ is not directly observable, the now quite familiar kernel estimates for the autoregression function are not applicable. We use instead nonparametric deconvolution estimators similar to those discussed in regression analysis by Fan and Truong (1993). These estimators are consistent and provide a convenient tool for exploratory data analysis helping in the decision which particular parametric model to choose for further analysis of the data.

2 A nonparametric stochastic volatility model

We consider some asset with price $S(t)$ at time t and, following Taylor (1994), define the return from an integer time $t-1$ to time t as

$$R_t = \log \frac{S(t)}{S(t-1)}.$$

To estimate a stochastic volatility model like (1.1) and (1.2), discretized versions of these equations are considered. Wiggins (1987) and Chesney and Scott (1989) use the Euler approximation

$$R_t = \mu + \sigma_{t-1} W_t \tag{2.1}$$

$$\log \sigma_t = \alpha + \phi\{\log \sigma_{t-1} - \alpha\} + \vartheta\, W_t^* \tag{2.2}$$

(W_t, W_t^*) denote i.i.d. bivariate standard normal random variables with zero mean and correlation ρ. In (2.1), the lagged quantity σ_{t-1} appears as the stochastic volatility for period t. This is rather advantageous for statistical purposes, as we will

clearly see later on.
As another simplification of (1.1), Taylor (1994) considers

$$R_t = \mu + \sigma_t W_t, \tag{2.3}$$

and he called (2.1), (2.2) a lagged autoregressive random variance (LARV) model, as $\log \sigma_t$ follows a linear autoregressive scheme. Analogously, (2.3), (2.2), together, is called a contemporaneous autoregressive random variance (CARV) model.

In this paper, we consider nonparametric generalizations of these models. We start with the lagged case and study it in detail, whereas we give a short discussion of the contemporaneous case at the end of Section 3.
We replace (2.2) by a nonlinear nonparametric model for $\xi_t = \log \sigma_t$:

$$\xi_t = m(\xi_{t-1}) + \eta_t, \tag{2.4}$$

where η_t denote i.i.d. zero-mean normal random variables with variance σ_η^2, and m is an arbitrary autoregression function for which we only require certain smoothness assumptions.
In order to ensure that the Markov chain (ξ_t) possesses nice probabilistic properties - e.g. geometric ergodicity and β-mixing (absolute regularity) or α-mixing (strongly mixing) with geometrically decaying mixing coefficients - it suffices (because of the assumption of normally distributed innovations η_t) to assume an appropriate drift condition on m, e.g.

$$\limsup_{|x|\to\infty} \left| \frac{m(x)}{x} \right| < 1, \tag{A1}$$

cf. Doukhan (1994), Proposition 6 (page 107). Then, in particular, ξ_t has a unique stationary distribution with density p_ξ.
We want to estimate m using kernel-type estimates. The usual Nadaraya-Watson estimates are, however, not applicable as we cannot observe the volatility σ_t or its logarithm ξ_t directly. The available data are the asset prices S_t or the returns R_t which are related to σ_t by (2.1). Taking logarithms and using the abbreviations

$$X_t = \frac{1}{2}\log(R_t - \mu)^2 - \mu_\varepsilon, \;\; \varepsilon_t = \frac{1}{2}\log W_t^2 - \mu_\varepsilon$$

with $\mu_\varepsilon = \mathcal{E}\left(\frac{1}{2}\log W_t^2\right) = -0.63518...$ (Scott (1987)), we get

$$X_t = \xi_{t-1} + \varepsilon_t, \tag{2.5}$$

where the ε_t are i.i.d. zero-mean random variables distributed as $\frac{1}{2}$ times the logarithm of a χ_1^2-random variable centered around 0. The correlation between the standard normal random variable W_t, appearing in the definition of ε_t, and η_t of (2.4) is ρ. (2.4), (2.5), together, form a nonparametric autoregressive model with errors-in-variables as ξ_t cannot be observed directly but is known only through its convolution with the i.i.d. random variables ε_t. Plugging (2.5) into (2.4) we obtain the following equation for X_t alone

$$X_t = m(X_{t-1} - \varepsilon_{t-1}) + \eta_{t-1} + \varepsilon_t. \tag{2.6}$$

Remark Assumption (A.1) also implies geometric ergodicity including geometrically β - and strong mixing for the process (X_t).

3 Kernel estimates for the autoregressive volatility function

Fan and Truong (1993) have studied a nonparametric regression model with errors-in-variables similar to the nonparametric autoregressive model (2.4), (2.5). Following their approach, we construct nonparametric estimates for m based on a sample $X_1, \ldots, X_T$. Let us assume that the parameter μ, which is the expectation of the returns R_t, is known such that the X_t are observable. From applications it can be justified that this expectation is close to zero. In case $\mu \neq 0$, the returns have to be centered before the procedure described below should be applied.
If we could observe $\xi_1, \ldots, \xi_T$ then we could estimate their stationary density, $p_\xi(x)$ by the kernel estimate

$$\hat{p}_\xi(x,h) = \frac{1}{Th} \sum_{t=1}^{T} K(\frac{x-\xi_t}{h}),$$

where K denotes a probability density and $h > 0$ denotes the bandwidth. The strongly mixing property of (ξ_t), which is ensured by (A1), immediately implies consistency via a covariance inequality.
As we only observe $X_1, \ldots, X_T$, whose stationary density is the convolution of p_ξ with the known density of the i.i.d. random variables ε_t, we have to use a deconvolution density estimate instead:

$$\hat{p}(x,h) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iwx} \phi_K(wh) \frac{\hat{\phi}_x(w)}{\phi_\varepsilon(w)} dw \tag{3.1}$$

with

$\phi_\varepsilon(w) = \mathcal{E}\ e^{iw\varepsilon_1}$, the characteristic function of ε_t,
$\phi_K(w) = \int_{-\infty}^{\infty} e^{iwx} K(x) dx$, the Fourier transform of the kernel K,
$\hat{\phi}_x(w) = \frac{1}{T} \sum_{t=1}^{T} e^{iwX_t}$, the sample characteristic function of $X_1, \ldots, X_T$.

The bandwidth h, depending on the sample size T, acts as a smoothing parameter as usual. For i.i.d. observations $\xi_1, \ldots, \xi_T$, the estimate $\hat{p}(x,h)$ for $p_\xi(x)$ has been investigated in detail by Stefanski and Carroll (1990), Carroll and Hall (1988), Fan (1991a,b) and Liu and Taylor (1989). Note that (3.1) can be written as a kernel estimator similar to $\hat{p}_\xi(x,h)$, namely

$$\hat{p}(x,h) = \frac{1}{Th} \sum_{t=1}^{T} K_h(\frac{x-X_t}{h})$$

with a kernel K_h depending on h and on the known distribution of the ε_t

$$K_h(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iwx} \frac{\phi_K(w)}{\phi_\varepsilon(w/h)} dw. \tag{3.2}$$

Remark. It should be noted, that without knowing anything of the distribution of the ε_t it is completely impossible to recover the stationary density p_ξ.

Now, the nonparametric estimate for $m(x)$ is defined as a Nadaraya-Watson estimate with kernel K_h and with X_t replacing ξ_t, more exactly

$$\hat{m}(x,h) = \frac{1}{Th} \cdot \sum_{t=1}^{T} K_h(\frac{x-X_t}{h})\, X_{t+1}/\hat{p}(x,h). \tag{3.3}$$

In order to apply this estimator it is necessary to evaluate the characteristic function ϕ_ε of ε_t and to make use of a kernel K for which the Fourier transform ϕ_K takes a convenient form. Concerning the explicit form and the asymptotic behaviour of ϕ_ε we have the following result.

Lemma 3.1: *Assume $W \sim \mathcal{N}(0,1)$, and let the density of the standard normal distribution be φ. The distribution of the centered random variable $\varepsilon = \frac{1}{2}\log W^2 - \mu_\varepsilon$ possesses the following density*

$$p_\varepsilon(x) = 2\ \varphi(e^{x+\mu_\varepsilon})\ e^{x+\mu_\varepsilon}, \quad x \in \mathbb{R}.$$

Here $\mu_\varepsilon = (\kappa + \log 2)/2 \approx 0.63518$ (κ denotes Eulers constant).
Let us denote by Γ the Gamma function. We have

$$\phi_\varepsilon(w) = \frac{e^{(\frac{\log 2}{2} - \mu_\varepsilon)iw}}{\sqrt{\pi}} \Gamma(\frac{1+iw}{2}), \quad w \in \mathbb{R}.$$

Concerning the tail behaviour of ϕ_ε we have for all d_0, d_1 with $0 < d_0 < \sqrt{2} < d_1 < \infty$:

$$d_0\ e^{-|w|\pi/4} \le |\phi_\varepsilon(w)| \le d_1 e^{-|w|\pi/4} \text{ as } |w| \longrightarrow \infty. \tag{3.4}$$

Proof: The explicit expressions for p_ε and ϕ_ε can be obtained by direct computation, while (3.4) is an immediate consequence of the tail-behaviour of Γ, which can be found for example in Gradstein and Ryshik (1981) (No. 8.328, page 331). ∎

Now, let us investigate the asymptotic behaviour of the kernel estimator $\hat{m}(\cdot, h)$, cf. (3.3). We have

$$\hat{m}(x,h) - m(x) = \frac{\frac{1}{Th} \sum_t K_h(\frac{x-X_t}{h})(X_{t+1} - m(x))}{\frac{1}{Th} \sum_t K_h(\frac{x-X_t}{h})}. \tag{3.5}$$

The following lemmas imply the consistency of $\hat{m}(\cdot, h)$.

Lemma 3.2: *Assume that m is twice continuously differentiable and that p_ξ is continuously differentiable. Assume that ϕ_K has a bounded support, $[-M_0, M_0]$ say, and that $h = h(T) = c/\log T$ where $c > M_0\pi/2$.*

(i) $$E \frac{1}{Th} \sum_t K_h(\tfrac{x-X_t}{h})(X_{t+1} - m(x)) = \int_{-\infty}^{\infty} \{m(u) - m(x)\} \frac{1}{h} K(\tfrac{x-u}{h}) p_\xi(u) du = O(h^2)$$

(ii) $$Var\big(\frac{1}{Th} \sum_t K_h(\tfrac{x-X_t}{h})(X_{t+1} - m(x))\big) = o(1).$$

Lemma 3.3: *Assume that p_ξ is twice times continuously differentiable. Assume that ϕ_K has a bounded support, $[-M_0, M_0]$, say, and that $h = h(T) \sim c/\log T$ where $c > M_0\pi/2$. Then*

(i) $$E \frac{1}{Th} \sum_t K_h(\tfrac{x-X_t}{h}) = \int_{-\infty}^{\infty} p_\xi(u) \frac{1}{h} K(\tfrac{x-u}{h})\, du = p_\xi(x) + O(h^2)$$

(ii) $$Var\big(\frac{1}{Th} \sum_t K_h(\tfrac{x-X_t}{h})\big) = o(1).$$

As an immediate consequence of Lemma 3.2 and 3.3 we obtain

Theorem 3.4: *Under the assumptions of Lemma 3.2 and 3.3 we obtain for all $x \in \mathbb{R}$*

$$(\log T)^2(\hat{m}(x,h) - m(x)) = O_p(1).$$

The nonparametric generalization of the contemporaneous autoregressive random variance model, where

$$X_t = \xi_t + \varepsilon_t \tag{3.6}$$

holds instead of (2.5), while the structure of (ξ_t) stated in (2.4) remains valid, is much more complicated to deal with. The problems arise from the fact that ξ_t and ε_t are not independent (as ξ_{t-1} and ε_t were before). To see this recall that ξ_t depends on η_t which itself is correlated to W_t (correlation ρ) appearing in the definition of ε_t. Thus, the stationary density of our observations X_t is for the contemporaneous case not the convolution of p_ξ (which we are interested in) with the known density of the i.i.d. random variables ε_t. To overcome the difficulties one could assume that $\rho = 0$ which together with the assumption of normality for the distribution of (η, W) implies independence even of ε_t and ξ_t. Under this assumption $\rho = 0$ all above results remain valid as can be easily seen.

In case we want to stay with the assumption $\rho \neq 0$ one has to look for another possibility to estimate p_ξ. One proposal may be as follows. Since

$$X_t = \xi_t + \varepsilon_t = m(\xi_{t-1}) + (\eta_t + \varepsilon_t)$$

we could estimate the characteristic function of $\mathcal{L}(m(\xi_o))$ by $\hat{\phi}_x(w)/\phi_{\eta+\varepsilon}(w)$. Here $\phi_{\eta+\varepsilon}$ denotes the characteristic function of the known distribution of $\eta_1 + \varepsilon_1$. Now $\xi_1 = m(\xi_0) + \eta_1$ which suggests the following deconvolution estimator for p_ξ

$$\begin{aligned} \tilde{p}(x,h) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iwx} \phi_K(wh) \frac{\hat{\phi}_x(w)}{\phi_{\eta+\varepsilon}(w)} \phi_\eta(w) dw \\ &= \frac{1}{Th} \sum_{t=1}^{T} \tilde{K}_h \left(\frac{x - X_t}{h}\right) \end{aligned}$$

where $\tilde{K}_h(n) = \dfrac{1}{2\pi} \displaystyle\int_{-\infty}^{\infty} e^{-iwn} \phi_K(w) \frac{\phi_\eta(w/h)}{\phi_{\eta+\varepsilon}(w/h)} dw.$

Finally, as a nonparametric estimator for m we propose

$$\tilde{m}(x,h) = \frac{1}{Th} \sum_{t=1}^{T} \tilde{K}_h\left(\frac{x - X_t}{h}\right) X_{t+1} / \tilde{p}(x,h).$$

We have, as before

Lemma 3.5: *Under suitable assumptions we have*

$$E\ \tilde{p}(x,h) = \int_{\mathbb{R}} p_\xi(x - hu) K(u) du = p_\xi(x) + O(h^2).$$

In order to obtain consistency of $\hat{p}(x,h)$ we computed above the variance and obtained that it converges to zero. For the proof (cf. proof of Lemma 3.3) it was rather essential to know the asymptotic behaviour of the characteristic function ϕ_ε appearing in the denominator of K_h. Similarily, we need for a consistency result for $\tilde{p}(x,h)$ some information on the asymptotic behaviour of $\phi_{\varepsilon+\eta}$, which seems to be a rather

delicate problem. A direct computation of $\phi_{\varepsilon+\eta}(w)$ leads to explicit expressions containing functions related to the so-called parabolic-cylinder functions $D_\nu(x)$. The argument w appears in the argument and in the parameter of D, and we were not able to quantify the asymptotic behaviour of such functions as $|w| \longrightarrow \infty$.
The same problems arise when dealing with the numerator of $\tilde{m}(x,h)$, For the numerator even the computation of its expectations does not lead to such nice expressions as in the lagged case.

Proofs.

Proof of Lemma 3.2:

(i) The expectation is equal to

$$\begin{aligned} & \frac{1}{h} E\ K_h(\frac{x-X_1}{h})(X_2 - m(x)) \\ = \ & \frac{1}{h} E\ K_h(\frac{x-\xi_0-\varepsilon_1}{h})(m(\xi_0)+\eta_1+\varepsilon_2 - m(x)) \\ = \ & \frac{1}{h} E\ K_h(\frac{x-\xi_0-\varepsilon_1}{h})(m(\xi_0) - m(x)) + \frac{1}{h} E\ K_h(\frac{x-\xi_0-\varepsilon_1}{h})\, \eta_1. \end{aligned}$$

Recall that $E\ \varepsilon_2 = 0$ and that ε_2 is independent of ξ_0 and ε_1. Unfortunately η_1 and ε_1 are not independent. But, because of the independence of ξ_0 and $(\varepsilon_1, \eta_1) = (\frac{1}{2}\log W_1^2 - \mu_\varepsilon, \eta_1)$ and $W_1 \sim \mathcal{N}(0,1)$

$$\begin{aligned} & E\ K_h(\frac{x-\xi_0-\varepsilon_1}{h})\, \eta_1 \\ = \ & \int_{\mathbb{R}^3} K_h(\frac{x-u-\frac{1}{2}\log w^2 + \mu_\varepsilon}{h})\ v\ p_\xi(u)\ p_{\eta|W=w}(v)\varphi(w) du\ dv\ dw \\ = \ & \int_{\mathbb{R}^2} K_h(\frac{x-u-\frac{1}{2}\log w^2 + \mu_\varepsilon}{h})\ \rho\sigma_\eta w\ p_\xi(u)\ \varphi(w)\ du\ dw \end{aligned}$$

since the conditional distribution of η given $W = w$ is $\mathcal{N}(\rho\sigma_\eta w,\ \sigma_\eta^2(1-\rho^2))$ by our assumptions. The latter integral is equal to zero by symmetry arguments (recall that the normal density φ is a symmetric function). Thus, the expectation under investigations equals

$$\begin{aligned} & \frac{1}{h} E\ K_h(\frac{x-\xi_0-\varepsilon_1}{h})(m(\xi_0) - m(x)) \\ = \ & \frac{1}{h}\int_{\mathbb{R}^2} K_h(\frac{x-u-v}{h})(m(u)-m(x))\ p_\xi(u)\ p_\varepsilon(v)\ du\ dv \\ = \ & \frac{1}{2\pi h}\int_{\mathbb{R}^3} e^{-i\frac{w}{h}(x-u-v)} \frac{\phi_K(w)}{\phi_\varepsilon(\frac{w}{h})}(m(u)-m(x))\ p_\xi(u)\ p_\varepsilon(v)\ du\ dv\ dw \\ = \ & \frac{1}{2\pi h}\int_{\mathbb{R}^2} e^{-i\frac{w}{h}(x-u)} \phi_K(w)(m(u)-m(x))\ p_\xi(u)\ du\ dw \\ = \ & \frac{1}{h}\int_{\mathbb{R}}\{\int_{\mathbb{R}} \frac{1}{2\pi}\, e^{-iw(x-u)/h}\phi_K(w)\ dw\}\,(m(u)-m(x))\ p_\xi(u)\ du. \end{aligned}$$

The expression in curly bracket is by Fourier inversion equal to $K((x-u)/h)$, thus we have proved the first part of (i).
A Taylor-expansion of m and p_ξ up to second (first) order yields because of $\int_{\mathbb{R}} v\ K(v)\ dv = 0$:

$$\begin{aligned}
& \int_{\mathbb{R}} \{m(u)-m(x)\}\ \frac{1}{h}\ K(\frac{x-u}{h})\ p_\xi(u)\ du \\
= & \int_{\mathbb{R}} K(v)\{-h\ v\ m'(x)+\frac{1}{2}h^2v^2m''(\hat{x}_1)\}\{p_\xi(x)-h\ v\ p'_\xi(\hat{x}_2)\}\,dv \\
= & \ O(h^2).
\end{aligned}$$

$\hat{x}_1$ and $\hat{x}_2$ denote suitable values between $x-h\ v$ and x, possibly depending on v.

(ii) Concerning the variance we obtain

$$\begin{aligned}
& \operatorname{var}\left(\frac{1}{Th}\ \sum_t\ K_h(\frac{x-X_t}{h})(X_{t+1}-m(x))\right) \\
= & \ \frac{1}{Th^2}\cdot\operatorname{var}\left(K_h(\frac{x-X_1}{h})(X_2-m(x))\right)+ \\
& +\frac{2}{T^2h^2}\cdot\sum_{s<t}\operatorname{cov}\left(K_h(\frac{x-X_s}{h})(X_{s+1}-m(x)),\ K_h(\frac{x-X_t}{h})(X_{t+1}-m(x))\right).
\end{aligned}$$

Using a covariance-inequality for strongly mixing sequences with geometrically decaying mixing coefficient (cf. Bosq (1996), Corollary 1.1 (page 19) we obtain the following bound of the above expression

$$\frac{1}{Th^2}\sup_{u\in\mathbb{R}}|K_h(u)|^2\ E(X_2-m(x))^2+\frac{O(1)}{Th^2}\left(E|K_h(\frac{x-X_1}{h})(X_2-m(x))|\right)^{\frac{2}{2+\delta}}$$

for $\delta>0$ arbitrarily small. Since

$$\left(E|K_h(\frac{x-X_1}{h})(X_2-m(x))|^{2+\delta}\right)^{\frac{2}{2+\delta}}\le\sup_{u\in\mathbb{R}}|K_h(u)|^2\cdot\left(E|X_2-m(x)|^{2+\delta}\right)^{\frac{2}{2+\delta}},$$

and since, from Fan and Truong (1993), (7.8), we have for $\chi=\frac{1}{4}M_0\pi>0$

$$\sup_{u\in\mathbb{R}}|K_h(u)|=O(h)+O(\frac{\exp(\chi/h)}{h}),$$

we can bound the variance through $O(\frac{\exp(2\chi/h)}{Th^4})$. This expression converges to zero for $h=c/\log T$ and $c>2\chi$.

Proof of Lemma 3.3:

(i) We have by independence of ξ_0 and ε_1

$$
\begin{aligned}
& E\ \frac{1}{h}\ K_h(\frac{x-X_1}{h}) \\
= & \ \frac{1}{h}\ E\ K_h(\frac{x-\xi_0-\varepsilon_1}{h}) \\
= & \ \frac{1}{h}\int_{\mathbb{R}^2} K_h(\frac{x-u-v}{h})\ p_\xi(u)\ p_\varepsilon(v)\, du\, dv \\
= & \ \frac{1}{2\pi h}\int_{\mathbb{R}^2} e^{-i\frac{w}{h}(x-u)}\phi_K(w)\ p_\xi(u)\, du\, dw \\
= & \ \frac{1}{h}\int_{\mathbb{R}} K(\frac{x-u}{h})\ p_\xi(u)\ du \\
= & \ \int_{\mathbb{R}} p_\xi(x-h\ v)\ K(v)\ dv = p_\xi(x) + O(h^2).
\end{aligned}
$$

The last equality is based on a second order Taylor-approximation of p_ξ.

(ii) Along the same lines as in the proof of Lemma 3.2 we obtain the wanted assertion.

Proof of Lemma 3.5:

$$
\begin{aligned}
& E\ \frac{1}{h}\ \tilde{K}_h(\frac{x-X_1}{h}) = \frac{1}{h}\ E\ \tilde{K}_h(\frac{x-m(\xi_0)-\varepsilon_1-\eta_1}{h}) \\
= & \ \frac{1}{2\pi h}\int_{\mathbb{R}^4} e^{-iw\ \frac{x-r-s-t}{h}}\phi_K(w)\, \frac{\phi_\eta(w/h)}{\phi_{\eta+\varepsilon}(w/h)}\ dP^{m(\xi_0)}(r) dP^{(\varepsilon,\eta)}(s,t)\ dw \\
= & \ \frac{1}{2\pi h}\int_{\mathbb{R}^2} e^{-iw\ \frac{x-r}{h}}\int_{\mathbb{R}^2} e^{iw\ \frac{s+t}{h}} dP^{(\varepsilon,\eta)}(s,t)\ dP^{m(\xi_0)}(r)\ \phi_K(w)\, \frac{\phi_\eta^{(w/h)}}{\phi_{\eta+\varepsilon}(w/h)}\ dw \\
= & \ \frac{1}{2\pi h}\int_{\mathbb{R}^2} e^{-iw\ \frac{x-r}{h}}\phi_K(w)\ \phi_\eta(w/h)\ dw\ dP^{m(\xi_0)}(r) \\
= & \ \frac{1}{h}\int_{\mathbb{R}^2} \frac{h}{\sigma_\eta}\varphi\left(\frac{\frac{x-r}{h}-u}{\sigma_\eta/h}\right)\ K(u)\ du\ dP^{m(\xi_0)}(r)
\end{aligned}
$$

because $\phi_K(w)\cdot\phi_\eta(w/h)$ is the characteristic function of $K * \mathcal{N}(0,\sigma_\eta^2/h^2)$ with density $\frac{h}{\sigma_\eta}\int_{\mathbb{R}}\varphi(\frac{\cdot - u}{\sigma_\eta/h})\ K(u)\ du$. ($\varphi$ denotes the density of the standard normal distribution) and the Fourier inversion formula.

$$= \int_{\mathbb{R}} \left\{ \frac{1}{\sigma_\eta} \int_{\mathbb{R}} \varphi(\frac{x-hu-r}{\sigma_\eta}) \; dP^{m(\xi_0)}(r) \right\} \; K(u)\; du$$

The term in curly brackets is the density of $\mathcal{L}(m(\xi_0)+\eta_1)$ which is p_ξ

$$= \int_{\mathbb{R}} p_\xi(x-hu)\; K(u)\; du = p_\xi(x) + O(h^2)$$

using the usual arguments and $\int_{\mathbb{R}} u\; K(u)\; du = 0$.

References

[1] Bosq, D. (1996). Nonparametric Statistics for Stochastic Processes. Lecture Notes in Statistics **110**, Springer-Verlag, New York.

[2] Carroll, R.J. and Hall, P. (1988). Optimal rates of convergence for deconvoluting a density. *J. Amer. Statist. Assoc.* **83**, 1184-1186.

[3] Chesney, M. and Scott, L.O. (1989). Pricing European Currency Options: A Comparison of the Modified Black-Scholes Model and a Random Variance Model. *J. Financial Quant. Anal.* **24**, 267-284.

[4] Doukhan, P. (1994). Mixing: Properties on Examples. Lecture Notes in Statistics **85**, Springer-Verlag, New York.

[5] Fan, J. (1991a). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19**, 1257-1272.

[6] Fan, J (1991b). Asymptotic normality for deconvolution kernel density estimators. *Sankhyā Ser. A* **53**, 97-110.

[7] Fan, J. and Truong, Y.K. (1993). Nonparametric regression with errors in variables. *Ann. Statist.* **21**, 1900-1925.

[8] Gradstein, I. and Ryshik, I. (1981). Tables. Verlag MiR, Moskau and Verlag Harri Deutsch, Thun.

[9] Liu, M.C. and Taylor, R.L. (1989). A consistent nonparametric density estimator for the deconvolution problem. *Canad. J. Statist.* **17**, 427-438.

[10] Scott, L.O. (1987). Option Pricing When the Variance Changes Randomly: Theory, Estimation and an Application. *J. Financial Quant. Anal.* **22**, 419-438.

[11] Scott, L.O. (1991). Random Variance Option Pricing: Empirical Tests of the Model and Delta-Sigma Hedging. *Adv. Futures Options Res.* **5**, 113-135.

[12] Stefanski, L.A. and Carroll, R.J. (1990). Deconvoluting kernel density estimators. *Statistics* **21**, 169-184.

[13] Taylor, S.J. (1994). Modelling Stochastic Volatility: A Review and Comparative Study. *Mathematical Finance* **4**, 183-204.

[14] Wiggins, J.B. (1987). Option Values Under Stochastic Volatility: Theory and Empirical Estimates. *J. Financial Econ.* **19**, 351-372.